Modelling of the GAL1 Genetic Circuit in Yeast Using Three Equations

Modelling of the GAL1 Genetic Circuit in Yeast Using Three Equations ⋆

Chi-Ching Hsu * **Yu-Heng Wu** * **Filippo Menolascina** ** **Torbjörn E. M. Nordling** *,***

* *Dept. of Mechanical Engineering, National Cheng Kung University, No. 1 University Road, Tainan 70101, Taiwan.*
** *School of Engineering, University of Edinburgh, Colin Maclaurin Road, EH9 3DW, Scotland, U.K.*
*** *Corresponding author (e-mail: torbjörn.nordling@nordlinglab.org).*

Abstract: Synthetic gene circuits can be used to modify and control existing biological processes and thus *e.g.* increase drug yields. Currently their use is hampered by the, largely, trial and error approach used to design them. Lack of reliable quantitative dynamical models of genetic circuits *e.g.* prevents the use of well established control design methods. We aim toward creation of a pipeline for automated closed-loop identification of dynamic models of synthetically engineered genetic circuits in microorganisms. As a step towards this aim, we here study modelling of a synthetically engineered GAL1 promoter circuit in S. cerevisiae that can be turned on by growing the yeast in Galactose and off by Glucose.
We perform parameter estimation on a system of three delay differential equations of Michaelis-Menten type based on *in vivo* data from a microfluidic experiment setup by Fiore *et al.* (2013). We show that three parameters are practically unidentifiable based on the data and simplify the model before fitting it to the data. The analysis and parameter estimation are done using AMIGO2–a state of the art MATLAB toolbox for iterative identification of dynamical models. We show that the goodness-of-fit of our model is superior to the four models proposed by Fiore *et al.*
Our plan is to use this model as a starting point for designing and conducting additional experiments in closed-loop until an approximation with desired predictive accuracy is obtained.

Keywords: synthetic biology, systems biology, genetic circuit, system identification, parameter estimation

1. INTRODUCTION

Established design methods with a field of control theory will be used to modify or control genetic circuits(????).

In the spirit of Norbert Wiener and Ludwig von Bertalanffy, we combine biology and engineering. The classical book on systems theory "General system theory"(?) and on cybernetics "Cybernetics or Control and Communication in the Animal and the Machine" (?) contain many biological examples and show that parts of theories were inspired by nature phenomena.

The importance of control theory in biological field is summarized in (?) by Vecchio *et al.*. Applying the knowledge of control theory in systems biology or synthetic biology, it is currently possible to modify or even control gene networks. We try to apply the control theory from engineering prospective on biological system. As Richard Feynman said, "What I cannot create, I do not understand." The final goal of synthetic biology is not only to modify or control genetic circuits, but also to create synthetic gene networks.

⋆ Chi-Ching Hsu, Yu-Heng Wu, and Torbjörn Nordling was in part supported by the Ministry of Science and Technology, Taiwan (grant 105-2218-E-006-016-MY2 and 105-2911-I-006-518), and the Innovation and Technology Program (I-Dream), National Applied Research Laboratories (NARLabs), Taiwan. Filippo Menolascina was in part supported by the Royal Society of Edinburgh, U.K.

1.1 Problem statement

This work aims toward designing an automated closed-loop parameter estimation of genetic circuits. Currently, synthetic genetic circuits are mainly designed using trial and error method, which is inefficient. As a result, we found out that model-based design become indispensable in synthetic genetic circuits. In addition, the key to model-based design is automated closed-loop parameter estimation of genetic circuits.

1.2 Genetic circuits

Fig. 1. GAL1 genetic circuits. The GFP protein was integrated downstream of the GAL1 promoter(?).

Strain of yeast cells with galactose gene (GAL1) is the system under investigation. We chose this genetic circuit since it has been biological well-analyzed. Moreover, Menolascina and coworkers have already conducted the *in vivo* experiment(**???**), so the real experimental data is available for us to do further analysis. As shown in Figure 1, the green fluorescent protein (GFP) was fused to GAL1 protein and expressed from the GAL1 promoter P_{GAL1}. The GAL1 protein utilize both galactose and glucose. However, The cells will consume all the available glucose first, since less energy is required. If glucose is available, the cells will never consume galactose.

Presence of galactose switchs on signal for the expression of the GAL1 gene. The activity of the GAL1 promoter P_{GAL1} is thus based on galactose(**?**). In short, the input could be either glucose (switch off GFP production) or galactose (switch on GFP production), yet not the combination of two, because the cells will consume only glucose rather than galactose if both sugars are available.

1.3 Experiment setting

Menolascina and coworkers applied an innovative experimental and computational method to do real-time automatic gene regulation. They controlled yeast cells in a series of works(**??**) with an innovative automated platform with microfluidic device, a time-lapse microscopy apparatus, and a set of syringes.

They tried to use control theory from engineering prospective to control the behavior of the GAL1 genetic circuit described in section 1.2. This platform was controlled entirely by computer according to the control algorithm. The experimental results were convincing in this work. In conclusion, *in-vivo*, real-time controlling gene networks is possible with this experiment setup.

Our project is based on these experimental setting and data. We try to do system identification, parameter estimation, and optimal experimental design on the GAL1 genetic circuit. In this report, our focus is on the system identification problem.

2. METHOD

This work follows the iterative identification procedure proposed by Balsa-Canto *et al.*(**?**). Fig. 2 represents the procedure for iterative identification. If there exists several model candidates, the left part in Figure 2 should be considered to find the most suitable model. The iterative identification procedure begins when the model framework is selected.

Our work focus on practical identifiability *a priori* and *a posteriori*. For the rest in Figure 2, see (**?**) for more detail.

2.1 Model fitting

After parameter estimation, the prediction output is plotted in Figure 3. The data we used is originally generated from Fiore *et al.*, which was published in 2013**?**.

We focus on several bumps present in the output GFP concentration. In Figure 3, we expected to see these bumps, yet we cannot observe clear tendency of slightly increasing or decreasing.

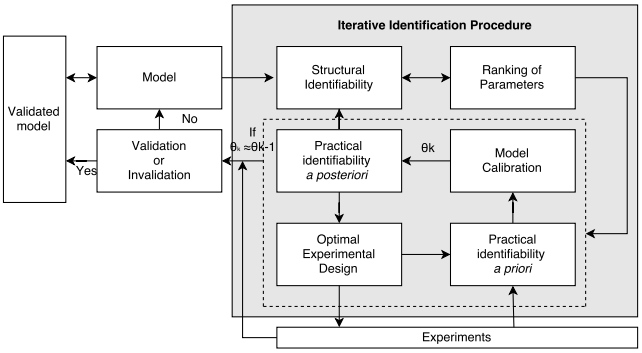

Fig. 2. Model building procedure(**?**).

2.2 Modify 3 equation ODE model

Based on the identifiability analysis and parameter estimation. We modified the structure of Gal1 no-delay model to the new 3 equation model. We added one more parameter α_2 and change previous α_2 to k_m.

$$\frac{dGAL1_{mRNA}}{dt} = \alpha_1 + V_{m_1}\frac{GAL1^{h_1}}{K_{m_1}^{h_1} + GAL1^{h_1}} - d_1 GAL1_{mRNA} \quad (1)$$

$$\frac{dGAL1_{fold}}{dt} = \alpha_2 + k_m GAL1_{mRNA} - d_3 GAL1_{fold} \quad (2)$$

$$\frac{dGAL1_{fluo}}{dt} = k_f GAL1_{fold} - k_b GAL1_{fluo} \quad (3)$$

To make it more concise. We set $d_2 - k_f$ equal to k_m.

3. CONCLUSION

Appendix A. MEASURES OF GOODNESS-OF-FIT

For the purpose of evaluate the fitting predictive ability of the 3 equation model and other models which will be mentioned in further chapter. RSS, R^2, adjusted R^2, FPE, AIC, and FIT was used to define as a comparison among different methods. FPE, AIC, and FIT was defined in Fiore *et al.* 2013**?**. The other definition is shown in follow.

$$\mathrm{RSS} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (A.1)$$

$$\mathrm{TSS} = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2 \quad (A.2)$$

$$\mathrm{ESS} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2 \quad (A.3)$$

Where y_i is model output for i-th data, $\hat{y}_i$ is the experiment measured output and $\bar{y}_i$ is the average value of measured output.

$$R^2 = ESS/TSS \quad (A.4)$$

$$aR^2 = 1 - (1 - R^2) * (\frac{n-1}{n-p-1}) \quad (A.5)$$

Where n is the sample size and p the amounts of model parameters.

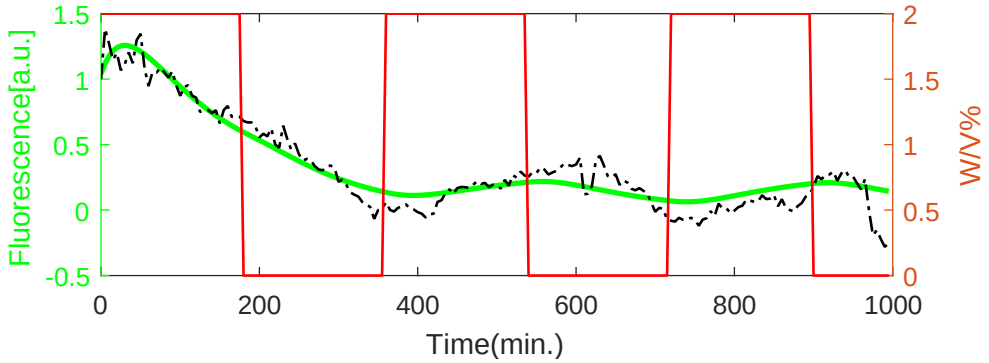

Fig. 3. Fitting a model after parameter estimation (green line) from *in vivo* experimental data (dash line). The red line is the galactose input. The quantity were measured in arbitrary units [a.u].

Appendix B. EXAMINATION OF FIORE ET AL. 2013

For the purpose of making sure Matlab toolbox AMIGO2 works and comparing the models in the paper with our 3 equation model. We reproduced the result of Fiore *et al.* 2013**?**. First of all, we used the experimental data providing by the paper and rebuilt the low pass filter to fit the experimental figure in the paper. The experimental data with an ideal transfer function type low pass filter $0.15/(s+0.15)$ is showing in figure [B.1]. Secondly we reproduced the ARX model, Transfer function model, State space model, and nonlinear model by using MATLAB control toolbox. The ideal square wave was treated as input to all the models to reproduce the result of scenario 2. But even through using the same parameters or the same strategy, the reproduced results are still vary different from the ones in the paper. We shown that in fig [B.2].

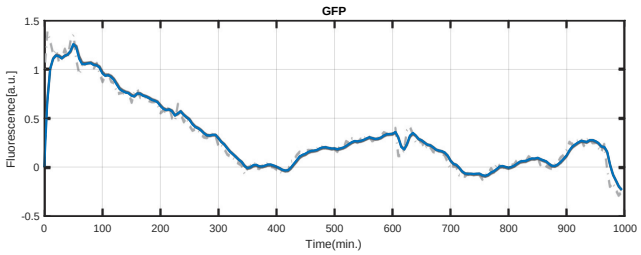

Fig. B.1. Low pass filter confirmation: Dash line and black line comes from Fiore *et al.* 2013**?** FIG. 3. Blue line is the Fluorescence level with low pass filter.

Even through the results was so different from the paper. We still calculate the validation of all the model. FPE, AIC, and FIT are all implemented in table B.1

B.1 Parameter estimation by using AMIGO2

We have implemented the parameter estimation by using AMIGO2 in our 3 equations model and transfer function model and the parameters are shown in following. For transfer function we set time delay as a flexible parameter.

- 3 equation parameters:
  $[\alpha,\ V_{m1},\ h_1,\ h_1,\ K_{m1},\ d_1,\ \alpha_2,\ d_2,\ K_f,\ K_b] = [9.9718\times10^{-7},\ 6.9265,\ 1.879,\ 3.5599,\ 8.5542\times10^{-2},\ 1.4196\times10^{04},\ 2.4395\times10^{-3},\ 2.6994\times10^{-2},\ 9.9472\times10^{-3}]$.
- Transfer function parameters:
  $[K_p,\ K_d,\ T_d] = [0.22599,\ 51.456,\ 30]$

The time delay problem in TF model is hard to implement in ODE solver. We have tried to use peda approximation to deal with it. I took Laplace transform of the ODE. and time delay term will become $U(s)e^(-Tds)$. Then I took Padé approximation to the $e^(-Td*s)$. It became $((-2/Td)+1)/((2/Td)+1)$. Finally, I calculated it and did the inverse Laplace. It returned the convolution form like $cos(f(t),U)$, $f(t)$ was a time-varying parameter. So I thought it won't work. I thought convolution won't work in the ODE solver

Table B.1. Value of the indices. arx: ARX model, tf: delayed transfer function, ss: state space model, nl: nonlinear model, tf-pe: transfer function after parameter estimation, and 3eq: 3 equation model.

	RSS	R^2	aR^2	FPE	AIC	FIT%
arx	34.26	3.56	3.58	0.17	-0.76	-13.63
tf	36.50	3.28	3.29	0.18	-0.73	-17.29
ss-pem	20.56	3.29	3.34	0.11	-0.97	11.98
nl	15.01	1.25	1.25	0.08	-1.12	24.77
tf-pe	29.55	4.53	4.70	0.16	-0.79	-5.53
3eq	2.65	1.21	1.22	0.01	-1.83	68.41

Appendix C. IFAC INSTRUCTIONS

This document is a template for LaTeX 2_ε. If you are reading a paper or PDF version of this document, please download the electronic file `ifacconf.tex`. You will also need the class file `ifacconf.cls`. Both files are available on the IFAC web site.

Please stick to the format defined by the `ifacconf` class, and do not change the margins or the general layout of the paper. It is especially important that you do not put any running header/footer or page number in the submitted paper. [1] Use *italics* for emphasis; do not underline.

Page limits may vary from conference to conference. Please observe the page limits of the event for which your paper is intended.

[1] This is the default for the provided class file.

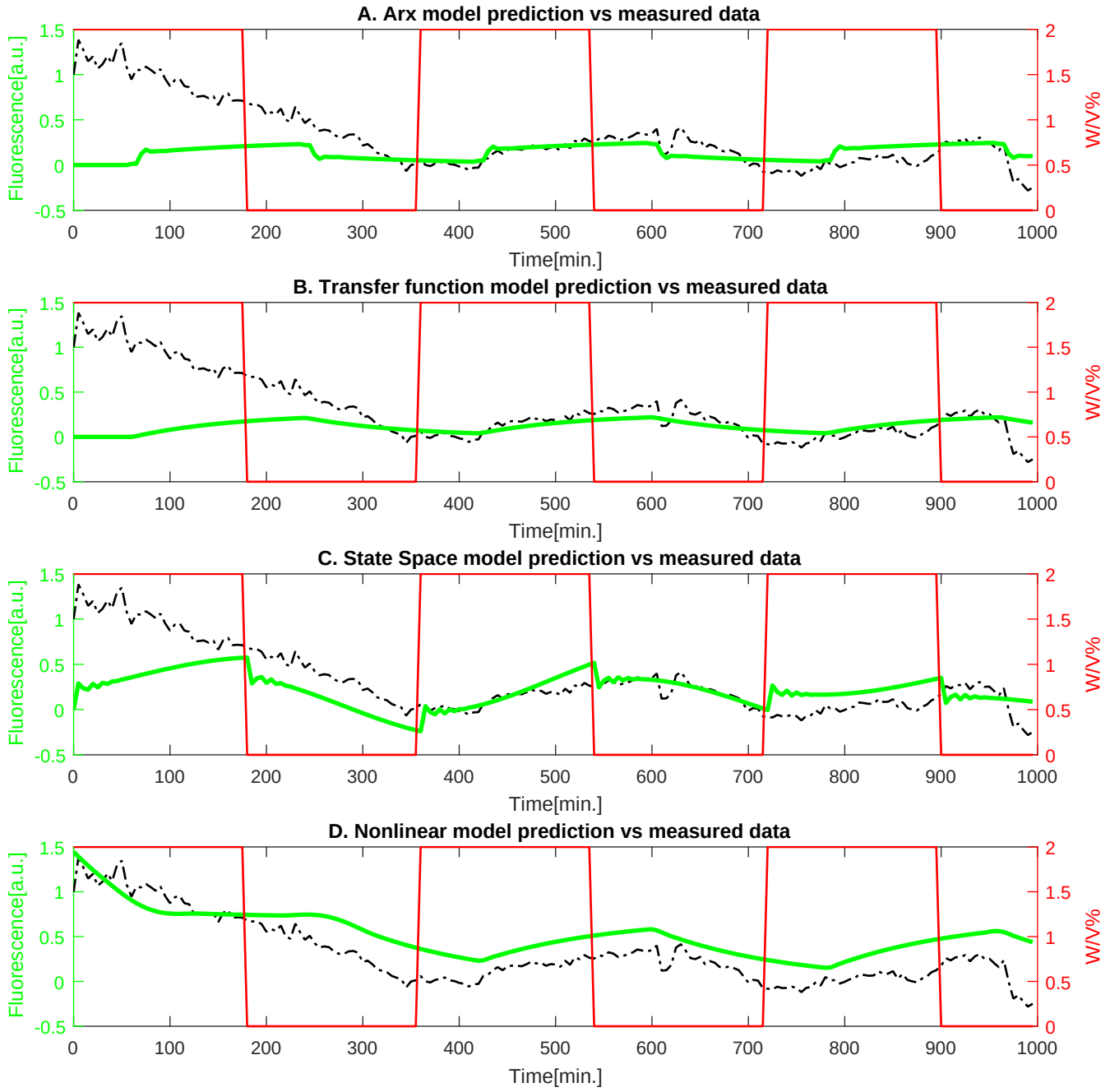

Fig. B.2. Scenario reproduced figure. First figure: The solid line represents the output of Arx model, the dashed line is the filtered experimental data. Second figure: The solid line represents the output of transfer function model, the dashed line is the filtered experimental data. Third figure: The solid line represents the output of state space model with PEM algorithms, the dashed line is the filtered experimental data.Forth figure: The solid line represents the output of nonlinear model, the dashed line is the filtered experimental data.Fifth figure: Galactose concentration is treated as an input.

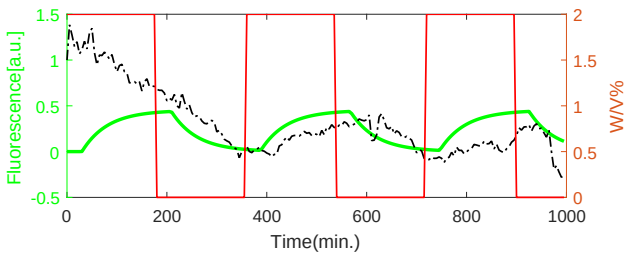

Fig. B.3. Model fitting after parameter estimation of transfer function model. Cell population average fluorescence (dashed line) treat as measured data. Solid line is the transfer function response.

Appendix D. PROCEDURE FOR PAPER SUBMISSION

Next we see a few subsections.

D.1 Review Stage

For submission guidelines, follow instructions on paper submission system as well as the event website.

Note that conferences impose strict page limits, so it will be better for you to prepare your initial submission in the camera ready layout so that you will have a good estimate for the paper length. Additionally, the effort required for final submission will be minimal.

D.2 Equations

Some words might be appropriate describing equation (D.1), if we had but time and space enough.

$$\frac{\partial F}{\partial t} = D\frac{\partial^2 F}{\partial x^2}. \tag{D.1}$$

See **?**, **?**, **?** and **?**.

Example. This equation goes far beyond the celebrated theorem ascribed to the great Pythagoras by his followers.

Theorem 1. The square of the length of the hypotenuse of a right triangle equals the sum of the squares of the lengths of the other two sides.

Proof. The square of the length of the hypotenuse of a right triangle equals the sum of the squares of the lengths of the other two sides.

Of course LaTeX manages equations through built-in macros. You may wish to use the `amstex` package for enhanced math capabilities.

D.3 Figures

To insert figures, use the `graphicx` package. Although other graphics packages can also be used, `graphicx` is simpler to use. See Fig. D.1 for an example.

Figures must be centered, and have a caption at the bottom.

D.4 Tables

Tables must be centered and have a caption above them, numbered with Arabic numerals. See table D.1 for an example.

Table D.1. Margin settings

Page	Top	Bottom	Left/Right
First	3.5	2.5	1.5
Rest	2.5	2.5	1.5

D.5 Final Stage

Authors are expected to mind the margins diligently. Papers need to be stamped with event data and paginated for inclusion in the proceedings. If your manuscript bleeds into margins, you will be required to resubmit and delay the proceedings preparation in the process.

Page margins. See table D.1 for the page margins specification. All dimensions are in *centimeters*.

D.6 PDF Creation

All fonts must be embedded/subsetted in the PDF file. Use one of the following tools to produce a good quality PDF file:

PDFLaTeX is a special version of LaTeX by Han The Thanh which produces PDF output directly using Type-1 fonts instead of the standard `dvi` file. It accepts figures in JPEG, PNG, and PDF formats, but not PostScript. Encapsulated PostScript figures can be converted to PDF with the `epstopdf` tool or with Adobe Acrobat Distiller.

Generating PDF from PostScript is the classical way of producing PDF files from LaTeX. The steps are:

(1) Produce a `dvi` file by running `latex` twice.
(2) Produce a PostScript (`ps`) file with `dvips`.
(3) Produce a PDF file with `ps2pdf` or Adobe Acrobat Distiller.

Appendix E. UNITS

Use SI as primary units. Other units may be used as secondary units (in parentheses). This applies to papers in data storage. For example, write "15 Gb/cm^2 (100 Gb/in^2)". An exception is when English units are used as identifiers in trade, such as "3.5 in disk drive". Avoid combining SI and other units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity in an equation. The SI unit for magnetic field strength **H** is A/m. However, if you wish to use units of T, either refer to magnetic flux density **B** or magnetic field strength symbolized as μ_0 **H**. Use the center dot to separate compound units, e.g., "A $\cdot$ m^2".

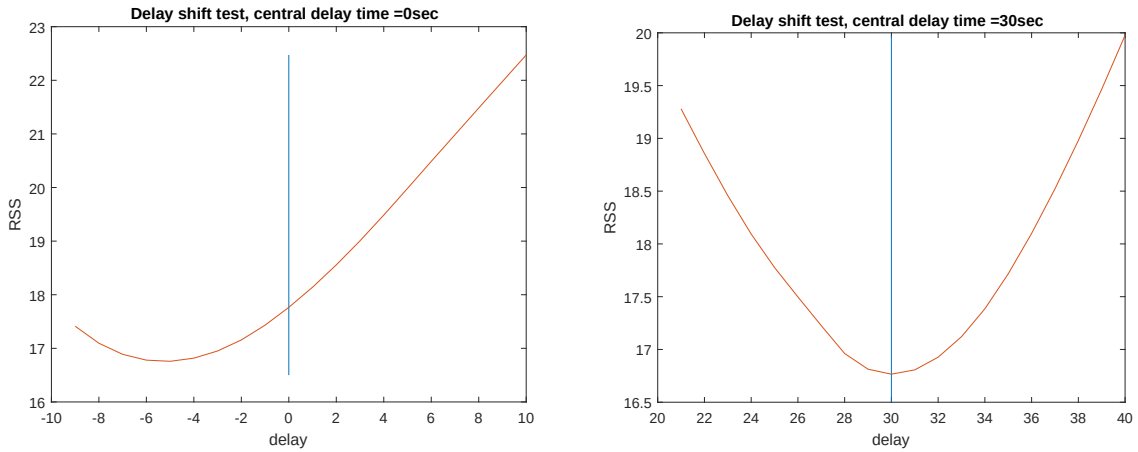

Fig. B.4. Delay shift test: We estimate the optimal delay time by calculating RSS value (red line) from central delay time (blue line) minus 10 to central delay time plus 10.

Fig. D.1. Bifurcation: Plot of local maxima of x with damping a decreasing

Appendix F. HELPFUL HINTS

F.1 Figures and Tables

Figure axis labels are often a source of confusion. Use words rather than symbols. As an example, write the quantity "Magnetization", or "Magnetization M", not just "M". Put units in parentheses. Do not label axes only with units. For example, write "Magnetization (A/m)" or "Magnetization (Am^{-1})", not just "A/m". Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)", not "Temperature/K".

Multipliers can be especially confusing. Write "Magnetization (kA/m)" or "Magnetization (10^3A/m)". Do not write "Magnetization (A/m) $\times$ 1000" because the reader would not know whether the axis label means 16000 A/m or 0.016 A/m.

F.2 References

Use Harvard style references (see at the end of this document). With LaTeX, you can process an external bibliography database using `bibtex`,[2] or insert it directly into the reference section. Footnotes should be avoided as far as possible. Please note that the references at the end of this document are in the preferred referencing style. Papers that have not been published should be cited as "unpublished". Capitalize only the first word in a paper title, except for proper nouns and element symbols.

F.3 Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have already been defined in the abstract. Abbreviations such as IFAC, SI, ac, and dc do not have to be defined. Abbreviations that incorporate periods should not have spaces: write "C.N.R.S.", not "C. N. R. S." Do not use abbreviations in the title unless they are unavoidable (for example, "IFAC" in the title of this article).

F.4 Equations

Number equations consecutively with equation numbers in parentheses flush with the right margin, as in (D.1). To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Use parentheses to avoid ambiguities in denominators. Punctuate equations when they are part of a sentence, as in

$$\int_0^{r_2} F(r,\varphi)drd\varphi = [\sigma r_2/(2\mu_0)] \cdot \int_0^{\inf} exp(-\lambda|z_j - z_i|)\lambda^{-1}J_1(\lambda r_2)J_0(\lambda r_i)d\lambda \tag{F.1}$$

Be sure that the symbols in your equation have been defined before the equation appears or immediately following. Italicize symbols (T might refer to temperature, but T is the unit tesla). Refer to "(D.1)", not "Eq. (D.1)" or "equation (D.1)", except at the beginning of a sentence: "Equation (D.1) is ...".

F.5 Other Recommendations

Use one space after periods and colons. Hyphenate complex modifiers: "zero-field-cooled magnetization". Avoid dangling participles, such as, "Using (1), the potential was calculated" (it is not clear who or what used (1)). Write instead: "The potential was calculated by using (1)", or "Using (1), we calculated the potential".

A parenthetical statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.) Avoid contractions; for example, write "do not" instead of "don' t". The serial comma is preferred: "A, B, and C" instead of "A, B and C".

[2] In this case you will also need the `ifacconf.bst` file, which is part of the `ifaconf` package.